Molecular Biology, Vol. 31, No. 4, 1997, pp. 616–622. Translated from Molekulyarnaya Biologiya, Vol. 31, No. 4, 1997, pp. 726–732.

UDC 577.113.3:57.087.1

Simulation of TATA Box Sequences in Eukaryotes

M. P. Ponomarenko, L. K. Savinkova, Yu. V. Ponomarenko, A. E. Kel', I. I. Titov, and N. A. Kolchanov

Institute of Cytology and Genetics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia

Received October 31, 1996

Abstract—We studied nucleotide sequences of TATA box in eukaryotic genes and their affinity to TATA-binding protein (TBP). The DNA affinity to TBP depended on the concentration of dinucleotides TA, TG, TC, AA, and AG. We proposed a method for predicting the DNA affinity to TBP based on its sequence, and applied it to promoters of eukaryotic genes. We observed the only peak of high affinity to TBP at position –30, which is the optimal TATA box position. Promoter sequences around the TATA box proved to have lower affinity to TBP than the arbitrary sequences. We simulated DNA sequences with high central and low terminal affinity to TBP, and showed reliable similarity between the simulation and the TATA boxes of eukaryotic genes.

Key words: TATA box, computer analysis, affinity, transcription

INTRODUCTION

It is common knowledge that TBP binding to the TATA box initiates formation of the transcriptional complex on the TATA-containing promoters [1–9]. TBP can bind a TATA box either alone or as a part of general transcriptional factor TFIID. Attachment of the general factors TFIIA, TFIIB, TFIIF, TFIIE, TFIIH, and RNA polymerase II to the TATA–TBP complex forms the general transcription complex [7–10]. Transcription complexes can be assembled in an alternative way on promoters with no TATA box. One of them is based on recognition of the initiating element (Inr) by the general factor TFIID or Inr-binding proteins. The Inr element has a conserved sequence YYANt/aYY bordering the starting nucleotide [11]. This element can operate autonomously [10].

TFIID was shown to be a protein complex consisting of TBP and additional protein factors (TAF) [12–14]. The molecular weight of TBP in man, mouse, and *Drosophila* is 40 kDa [15, 16], while in yeast it is 27 kDa [16–21]. It features a variable N-terminal and a conserved C-terminal domains [17–22]. Eighty-percent identity is specific for amino acid sequences of the TBP conserved domain in man, mouse, *Drosophila*, *Arabidopsis*, and yeast [16]. This conserved domain is necessary and sufficient for DNA binding, maintaining the basic level of transcription, and interaction with various transcription factors [12, 22–24]. At the same time, the variable TBP domain can be involved in specific protein–protein interactions required for assembly of the transcription complex on TATA-containing promoters.

X-ray analysis has shown that yeast TBP consists of four α-helices and a saddle β-sheet with the concave surface embracing the minor groove [25–27]. Although numerous publications have demonstrated that TBP binds to DNA as a monomer, gel filtration, affinity chromatography, and chemical cross-linking methods indicate that free TBP is a dimer [28].

It is known that the TBP affinity to TATA box affects the promoter efficiency both *in vivo* and *in vitro* [29]. The subsequent assembly of other general factors of RNA polymerase II in a functional transcription complex also depends on the initial interaction between TBP and TATA box.

By contrast to the majority of control proteins [30], TBP interacts with the minor groove of DNA [25–27]. The problem of recognizing nucleotide substitutions in the TATA box by TBP still remains unanswered. The equilibrium dissociation constants for complexes of the yeast TBP with TATA-containing promoter fragments (300 bp), which reflect TBP affinity to the DNA, fall within 2–3 nM [31–34]. Substitutions G ⟶ C and C ⟶ G in the TATA box significantly affect TBP affinity to the DNA, supposedly, through steric obstacles to TBP binding to the guanine extracyclic 2-amino group [35]. Still the background of these changes in TBP affinity to DNA induced by

0026-8933/97/3104-0616$18.00 © 1997 МАИК Наука /Interperiodica Publishing

A ⟶ T and T ⟶ A substitutions remains unclear, despite 10–100-fold changes in the equilibrium dissociation constant of the TBP–DNA complex [31] and 10–50-fold changes in its half-life [36].

Electrophoretic methods indicate [36] correlation between the stability of the TBP–TATA complex and the extent of DNA bending at the binding site. The more pronounced is DNA bending, the longer is the half-life of such complexes and the higher is the induced transcription.

The changes in TBP binding to the DNA fragments with and without TATA box were studied to understand the mechanism of TBP binding the class I, II, and III promoters [35]. According to the kinetic indices of TBP–DNA binding, the mechanism of binding is common for TATA and non-TATA DNA fragments. The proposed model suggests that free TBP dimer dissociates, and each of the monomers can form unstable complexes with DNA and slide along it until recognizing a TATA box. In this case a stable TBP–TATA complex is formed, which initiates the above-mentioned assembly of the general transcription complex [35].

However, despite abundant data on the TBP–TATA interaction, specific features of the TATA box responsible for its affinity to TBP are yet not recognized [37, 38].

Here we studied the nucleotide sequences of TATA boxes and revealed the features related to the affinity to TBP. We have developed a method for predicting a DNA affinity to TBP and used it to analyze nucleotide sequences of eukaryotic promoters. There appeared only one peak of high affinity to TBP at the –30 position, which is optimal for the TATA box. The sequences neighboring the TATA box have lower affinity to TBP than arbitrary sequences. We simulated DNA sequences with high central and low terminal affinity to TBP. Reliable similarity between such model DNA sequences and the natural eukaryotic TATA boxes was demonstrated.

MATERIALS AND METHODS

We used the data on TBP–DNA affinity presented in Table 1 [39] for TBP isolated from *Saccharomyces cerevisiae* (it is a common knowledge that yeast and human TBP are interchangeable in *in vivo* and *in vitro* systems). Sokolenko *et al.* [39] measured equilibrium dissociation constants K_d of TBP complex with 15-bp DNA fragments. Negative logarithms of the TBP–DNA complex dissociation constants ($-\ln[K_d]$) were used as a measure of TBP affinity to DNA (Table 1).

We studied 19 DNA fragments (Table 1) of human H1 histone and β-actin genes promoters. They differed in both A, T, G, and C content and the position of TATA-like sequences (Table 1). Two DNA fragments with no TATA box were also included.

Table 1. TBP affinity to 15-bp DNA fragments [39]

No.	Sequence	Affinity, $-\ln[K_d]$
1	CGCCCTATAAAACCC	24.23
2	GGGTTTTATAGGGCG	24.23
3	GCGCCCTATACTACC	24.08
4	GGTAGTATAGGGCGC	24.08
5	GCGCCCTGTACTACC	21.50
6	GGTAGTACAGGGCGC	21.50
7	GCGCCCTATGCTACC	21.17
8	GGTAGCATAGGGCGC	21.17
9	CGGCTTATATAAGCC	21.06
10	GGCTTATATAAGCCG	21.06
11	CGCCCAAACCCTATA	20.72
12	TATAGGGTTTGGGCG	20.72
13	TATAAAACCCAGCGG	17.73
14	CCGCTGGGTTTTATA	17.73
15	GTTTTTTTTTTCGCG	16.60
16	GGTGGCGCACGCCTG	16.27
17	AGAGTTCAAGACGAT	15.76
18	CATGGCGGCGGGGCG	11.87
19	CCCCCCCCCCCCCCC	11.78

Table 1 shows that the TBP–DNA affinity ($-\ln[K_d]$) varied from 11.78 to 24.23, which corresponds to the difference in the dissociation constant K_d by five orders of magnitude. This indicates significant influence of the nucleotide sequence on its affinity to TBP. Still we do not know what features of the DNA are responsible for this affinity.

Note that the traditional consensus and weight matrix methods of DNA sequence analysis [40] cannot be used for studying the DNA fragments presented in Table 1, since the are only applicable to highly similar sequences. Hence, these data [39] were analyzed with the help of SITEVIDEO computer system developed earlier for studying nucleotide sequences with low similarity [41, 42].

This system analyzes the number and position of a given dinucleotide z_1z_2 in each 15-nucleotide sequence $S = s_1 \ldots s_i \ldots s_{15}$. Each dinucleotide $z_1z_2 = \beta\gamma$ is designated by IUPAC codes. Weighed concentration of $\beta\gamma$ dinucleotide is calculated for a sequence S:

$$X_{\beta\gamma w}(S) = \sum_{i=1}^{14} \delta_{\beta\gamma}(s_i s_{i+1}) w(i), \qquad (1)$$

where $\delta_{\beta\gamma}(s_i s_{i+1}) = 1$ for $\beta = s_i$ and $\gamma = s_{i+1}$ or $\delta_{\beta\gamma}(s_i s_{i+1}) = 0$ in all other cases, $w(i)$ is weight function at position i.

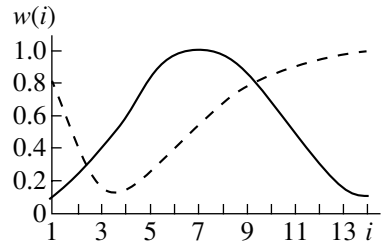

Fig. 1. Weight functions simulating the highest central (solid line) and terminal (dotted line) contribution of the dinucleotides to TBP affinity to the DNA.

The weight function $w(i)$ was introduced in equation (1) to account for the influence of the $\beta\gamma = s_i s_{i+1}$ dinucleotide at position i on the DNA affinity to TBP. The more is $w(i)$ value, the more is contribution of the dinucleotide $s_i s_{i+1}$ to the TBP–DNA affinity. Figure 1 presents two examples of the weight function $w(i)$ used in this work. Solid curve presents the weight function $w(i)$ of the dinucleotides $\beta\gamma$ with the highest contribution to the TBP–DNA affinity at the center of the DNA fragment, while dotted curve corresponds to the weight function of the dinucleotides with the terminal contribution to the TBP–DNA affinity. The number of such weight functions $w(i)$ totals 180 with different position and size of the DNA regions with the highest weights.

Since we know neither the dinucleotides $\beta\gamma$ responsible for the TBP–DNA affinity nor their position in the DNA, all $15 \times 15 = 225$ possible $\beta\gamma$ dinucleotides are analyzed, each of them is applied to all 180 weight functions $w(i)$. Hence, we consider all $225 \times 180 = 40{,}500$ concentrations $X_{\beta\gamma w}$ calculated by equation (1).

First, each of these concentrations is analyzed independently. For given nucleotides β, and γ, and a weight function $w(i)$ the concentration $X_{\beta\gamma w}(S)$ is calculated by equation (1) for all nucleotide sequences S. As a result we know concentrations $X_{\beta\gamma w}(S)$ for all DNA fragments with known affinity to TBP ($-\ln[K_d]$).

Correlation between these values is tested using the linear correlation index [43]:

$$r(-\ln[K_d]; X_{\beta\gamma w}) > R_{v=n-2;\alpha=0.05}, \quad (2)$$

where $R_{v;\alpha}$ is the critical r at the significance level α for n considered nucleotide sequences.

The condition (2) is tested for each of 40,500 $X_{\beta\gamma w}$ concentrations. If a given $X_{\beta\gamma w}$ concentration meets this condition, the relationship between the TBP–DNA affinity and this concentration is considered as reliable and such concentration is selected for further analysis. Otherwise the TBP–DNA affinity does not depend on concentration $X_{\beta\gamma w}$ and is omitted. This results in a list of concentrations $X_{\beta\gamma w}$ that determine DNA affinity to TBP.

Then all concentrations $X_{\beta\gamma w}$ satisfying condition (2) are analyzed to reveal interrelated concentrations with arbitrary nucleotide sequences (25 15-bp sequences with equal nucleotide frequencies). Each concentration $X_{\beta\gamma w}$ selected by condition (2) is calculated using equation (1). Similarly to the condition (2), correlation between two concentrations $X_{\beta\gamma w}$ and $X_{\phi\mu v}$ in these arbitrary sequences is also tested by the linear correlation index:

$$r(X_{\beta\gamma w}, X_{\phi\mu v}) > R_{v;\alpha=0.05}. \quad (3)$$

Condition (3) is tested for all pairs of the selected concentration. If it is not satisfied for a given pair of concentrations $X_{\beta\gamma w}$ and $X_{\phi\mu v}$, they are considered as representing different relationships between the TBP–DNA affinity and its nucleotide sequence and both are left in the list of selected concentrations. When the condition (3) is satisfied, the concentrations $X_{\beta\gamma w}$ and $X_{\phi\mu v}$ are considered interrelated for arbitrary DNA. In this case they represent the same relationship between the TBP–DNA affinity and its nucleotide sequence and the concentration with lower linear correlation index is removed from the list of selected concentrations:

$$r(X_{\beta\gamma w}; -\ln[K_d]) > r(X_{\phi\mu v}; -\ln[K_d]). \quad (4)$$

Thus, only independent concentrations representing different types of relationship between the TBP–DNA affinity and its nucleotide sequence remain in the list of selected concentrations $X_{\beta\gamma w}$.

RESULTS AND DISCUSSION

DNA sequence properties responsible for its affinity to TBP were studied using eight out of 19 DNA fragments presented in Table 1 (3–6, 11, 12, 15, and 19). These fragment served as a training sample for the analysis, while the remaining 11 fragments (1, 2, 7–10, 13, 14, and 16–18) were not analyzed which made possible their use as "independent data" to assess the reliability of analysis.

Analysis of the training sample using equations (1–4) has revealed two dinucleotides TV and WR whose weighed concentrations correlate with the TBP–DNA affinity. One of them, TV = {TA, TG, TC}, has the contribution to the affinity growing with closeness to the DNA fragment center (Fig. 1, solid curve). The highest index of linear correlation to the TBP–DNA affinity ($r = 0.765$ and $\alpha < 0.05$) was observed for the weighed concentration X_{TVw}. It is a common knowledge that subsequence TATA is the most conserved in the TATA box which was named after it. Dinucleotides TA composing this subsequence is a special case of the TV dinucleotide. Interestingly, subsequences TATG and TATC composed by the TV dinucleotide are also common in eukaryotic TATA boxes.

Another dinucleotide, WR = {AA, TA, AG, TG} has the contribution to the TBP–DNA affinity growing with closeness to the DNA fragment termini (Fig. 1, dotted curve). The linear correlation between the weighed concentration of the dinucleotide X_{WRw} and TBP–DNA affinity was reliable ($r = 0.728$ and $\alpha < 0.05$) for the training sample. At the same time there was no correlation between the concentrations X_{WRw} and X_{TVw} ($r = 0.04$ and $\alpha > 0.25$) in the sample of 25 arbitrary sequences.

The obtained results indicate that the dinucleotides WR at the DNA fragment termini also make significant contribution in its affinity to TBP. TA is a special case of the WR dinucleotide. Clearly, terminal TATA subsequence should affect affinity of the DNA fragment to TBP.

Prediction of TBP affinity to DNA by its nucleotide sequence. We generated a linear regression relationship between the TBP–DNA affinity and weighed concentrations of dinucleotides TV and WR for arbitrary 15-nucleotide sequence S:

$$-\ln[K_d(S)] = a_0 + a_1 X_{TVw}(S) + a_2 X_{WRv}(S), \qquad (5)$$

where $a_0 = 14.53$, $a_1 = 2.53$, and $a_2 = 0.87$ are coefficients of linear regression.

Equation (5) was obtained on the training sample with the above eight DNA fragments. It was tested on the "independent" sample of 11 DNA fragments used neither for revealing the X_{TVw} and X_{WRw} concentrations nor for generation of equation (5).

Figure 2 visualizes testing of equation (5) with "independent" sample. DNA affinity to TBP predicted by the equation (5) reliably corresponded to the experimental values [39] ($r = 0.785$ and $\alpha < 0.05$).

Hence, the equation (5) reliably predicts the DNA affinity to TBP by its nucleotide sequence.

TBP affinity to eukaryotic gene promoters. Equation (5) was used to analyze TBP affinity to eukaryotic gene promoters. We used 776 promoters from EPD database (release 45). We considered the nucleotide sequences from –257 to +108 relative to the transcription start. We calculated the affinity to TBP of each 15-nucleotide region of these sequences with the center at position i using equation (5) $-\ln[K_d(s_{i-7}\ldots s_i\ldots s_{i+7})]$ and plotted TBP affinity to the promoter against position i ($-250 \le i \le 101$). Such curves were obtained for all 776 promoters.

Figure 3 presents the resulting curve averaged for all promoters (bold line). Normal line shows an analogous relationship for 10,000 arbitrary sequences of a similar length with even nucleotide frequencies. By contrast to the promoters, one can see no correlation between the TBP–DNA affinity and the position number. Promoters have sole peak of TBP affinity with the top value $-\ln[K_d] = 22.26$ at position –30. The peak position coincides with the known optimal position of the TATA box, which agrees with the common view of the TATA box as the only TBP binding site.

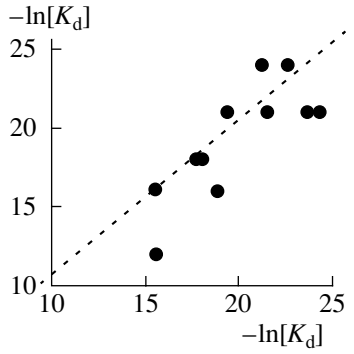

Fig. 2. Predicted (abscissa) and experimental (ordinate) TBP affinity to the DNA for independent data. Coefficient of linear correlation $r = 0.785$ ($\alpha < 0.05$).

In addition, the promoter sequences bordering TATA box have lower affinity to TBP as compared to arbitrary sequences. Note that the regions of reduced affinity to TBP are quite long (Fig. 3). The regions of the affinity upstream and downstream of the TATA box are 171 (positions –210 to –40) and 80 bp (position –20 to +61), respectively. We propose that these promoter regions provide for fine positioning of the TBP-TATA complex relative to the transcription start.

One can propose a kinetic interpretation of low promoter affinity to TBP at long regions bordering the TATA box. In essence, under certain conditions TBP starts sliding along these DNA regions towards the TATA box. Evaluations of the kinetic indices for such sliding will be published separately.

Simulation of eukaryotic TATA box sequences. Presence of promoter regions with low affinity to TBP around the TATA box was tested by simulating DNA sequences with high central and low terminal affinity to TBP. Arbitrary 35-nucleotide sequences with even nucleotide frequencies were considered. TBP affinity to each sequence was calculated by the equation (5) for three 15-nucleotide fragments: left (positions 1–15, S_{1-15}), central (positions 11–25, S_{11-25}), and right (positions 21–35, S_{21-35}). We tested the condition that TBP to the left and right regions is lower as compared to the central one by Δ:

$$\begin{aligned} &\mathrm{MAX}\{-\ln[K_d(S_{1-15}), -\ln[K_d(S_{21-35})\} \\ &\quad < -\ln[K_d(S_{11-25})] - \Delta, \end{aligned} \qquad (6)$$

where Δ is the selection threshold that equaled 6 in our simulation.

Using condition (6), we simulated lower TBP affinity to promoter DNA around the TATA box (Fig. 3). Only 10^3 out of 10^8 tested arbitrary sequences satisfied condition (6).

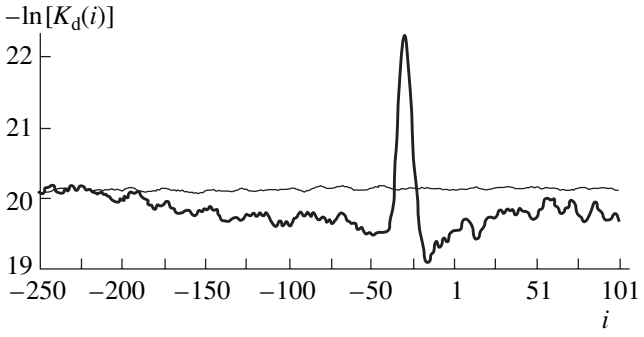

Fig. 3. Profile of TBP affinity to DNA along arbitrary sequences (normal line) and eukaryotic promoters (bold line).

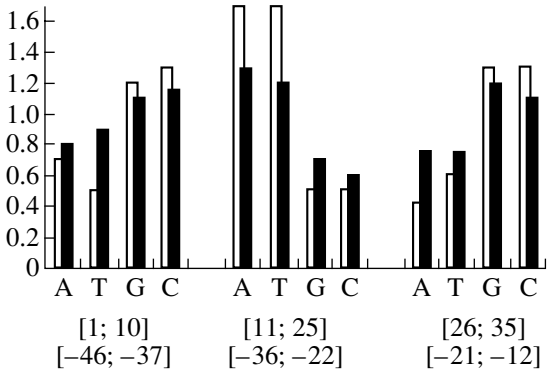

Fig. 4. Relative frequencies of nucleotides in promoters (solid bars) and simulation DNA sequences (empty bar). Coefficient of linear correlation $r = 0.91$ ($\alpha < 0.01$).

We selected 100 DNA sequences out of this group with the highest affinity to TBP at the center. Thus we simulated peak of high promoter affinity to TBP at position –30 (Fig. 3).

Table 2 presents best 10 out of the 100 simulation sequences and the consensus generated from all 100 sequences with reliably conserved nucleotides ($\alpha < 0.01$). The last line of Table 2 presents consensus sequence of the natural TATA boxes [40]. One can note good correspondence between consensus of the simulation sequences selected by affinity to TBP and TATA boxes (13 matches out of 15).

To test reliability of similarity between the simulation and promoter DNA sequences we first analyzed the nucleotides frequencies. We selected three regions in the promoter sequences of eukaryotic genes (–46…–37, –36…–22, and –21…–12) so that the optimal TATA box resided in the center of the second region. We calculated the nucleotide frequencies in each of these regions using the above mentioned 776 promoters from the EPD (release 45). The influence of odd heterogeneity in the DNA sequences that changes the observed nucleotide frequencies from the real ones was reduced by relating the nucleotide frequency at each region to the mean value for all regions.

In the simulation DNA sequences the corresponding three regions were 1–10, 11–25, and 26–35. Relative nucleotide frequencies were calculated for each of these regions.

The obtained results are shown in Fig. 4. The central region in both simulation and promoter sequences is AT-rich, while the left and right regions are AT-poor. In general the coefficient of linear correlation between the nucleotide frequencies in the simulation and promoter sequences $r = 0.91$ ($\alpha < 0.01$), which indicates reliable similarity between these sequences.

We carried out an additional test of similarity between the simulation and promoter sequences by the method for recognizing TATA boxes using weight

Table 2. Simulation DNA sequences with high central and low terminal affinity to TBP

No.	Central affinity	Sequence
1	26.74	GGAAGCGCCGCTCATATATATATGGCCCGAGCCAC
2	26.66	CAGAAAAACCGTCTTATATAAAAGAAGGCCCTTCA
3	26.00	CGATGGCCGCCCATATATATACAGGCAGCCCGGTG
4	26.00	CGACGCGCCCCCATATATATATGGGGGCAATGGTG
5	25.83	GGGCCCGAAGTGCTTATATATAAGCCCGGCGACGT
6	25.73	CGCCGGGCCGGCCTTATATATGAGGGCGTTTTCAC
7	25.54	CAGAAGCCGCGTCGTTCTATAGAAGGCCGCGGGGG
8	25.14	GAACCCGCGCCCCTATCTTAGAAAGGCGGCGCTCG
9	24.79	AAGCCGAACGGGCTATATACTGAGCGCCGGGGGAT
10	22.73	TCCCCGGCGGCTTTGTATAACAGCAGCCCCGCGTG
Consensus	Simulation DNA	−−−C−GSSSSCCCTTTWWWWAAAGSSSSSSSC−−G
	TATA box [40]	STWTAWADRSSSSSS

matrix [40]. We analyzed each of 776 nucleotide sequences of eukaryotic promoters from the EPD (release 45). TATA boxes were found in 71% promoters. In a similar way we analyzed each of the 100 simulation DNA sequences. In this case TATA boxes were found in 67% simulation sequences. The exact Fisher test indicates that the simulation and promoter sequences do not differ in the TATA box recognition frequencies.

CONCLUSION

The results obtained indicate that the DNA affinity to TBP is defined by its nucleotide sequence. This affinity increases with concentration of dinucleotides TV = {TA, TG, TC} in the center of the TBP–DNA binding site and dinucleotides WR = {AA, AG, TA, TG} at the termini of this site. Due to this relationship, eukaryotic promoters feature sole peak of high affinity to TBP at position –30, which is also the optimal TATA box position (Fig. 3). Moreover, this peak is surrounded by quite long regions of promoter DNA with affinity to TBP lower even than in arbitrary sequences. This leads to the conclusion that eukaryotic promoters are optimized for exact positioning of the TBP–TATA complex on the transcription start nucleotide. Simulation of the nucleotide sequences with high central and low terminal affinity to TBP demonstrate their reliable similarity to the natural promoters of eukaryotic genes.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 95-04-12469 and 97-04-49740), State program Human Genome, and Russian Department of Science, Education, and Technical Politics.

REFERENCES

1. Nussinov, R., *Crit. Rev. Biochem. Mol. Biol.*, 1990, vol. 25, pp. 185–224.
2. Mitchell, P.J. and Tjian, R., *Science*, 1989, vol. 215, pp. 371–378.
3. Iyer, V. and Struhl, K., *Mol. Cell. Biol.*, 1995, vol. 15, pp. 7059–7066.
4. Reinberg, D. and Roeder R., *J. Biol. Chem.*, 1987, vol. 262, pp. 3310–3321.
5. Woblbe, C.R. and Struhl, K., *Mol. Cell. Biol.*, 1990, vol. 10, pp. 3859–3867.
6. Reinberg, D., Horikoshi, M., and Roeder, R., *J. Biol. Chem.*, 1987, vol. 262, pp. 3322–3330.
7. Buratowski, S., Hahn, S., Guarente, L., and Sharp, P., *Cell*, 1989, vol. 56, pp. 549–561.
8. Sawadogo, M. and Sentenac, A., *Annu. Rev. Biochem.*, 1990, vol. 59, pp. 711–754.
9. Littie, J.W. and Green, M.K., *Nature*, 1989, vol. 57, pp. 103–113.
10. Buratowski, S., *Cell*, 1994, vol. 77, pp. 1–3.
11. Wingender, E., *Gene Regulation in Eukaryotes*, New York: VCH Weinheim, 1993, pp. 75–84.
12. Conaway, R. and Conaway, V., *Annu. Rev. Biochem.*, 1993, vol. 62, pp. 161–190.
13. Smale, S., Schidt, M., and Berk, A., *Proc. Natl. Acad. Sci. USA*, 1990, vol. 87, pp. 4509–4513.
14. Pugh, B.F. and Tjian R., *Cell*, 1990, vol. 61, pp. 1187–1197.
15. Smale, S.T. and Baltimore D., *Cell*, 1989, vol. 57, pp. 103–113.
16. Hoffman, A., Sihn, E., and Roeder R.G., *Nature*, 1990, vol. 346, pp. 387–390.
17. Horikoshi, M., Yamamoto, T., and Roeder, R.G., *Cell*, 1990, vol. 61, pp. 1171–1178.
18. Horikoshi, M., Wang, C., Fujii, H., Cromlish, J.A., Weil, P.A., and Roeder, R.G., *Nature*, 1989, vol. 341, pp. 299–301.
19. Peterson, M., Tanese, N., Pugh, B., and Tjian, R., *Science*, 1990, vol. 248, pp. 1625–1631.
20. Gashch, A. and Hoffman, A., *Nature*, 1990, vol. 340, pp. 390–393.
21. Lieberman, P., Schmidt, M., Kao, C., and Berk, A., *Mol. Cell. Biol.*, 1991, vol. 11, pp. 63–74.
22. Meisterernst, M. and Roeder, R., *Cell*, 1991, vol. 67, pp. 557–567.
23. Wiley, S.R., Kraus, R.J., and Mertz, J., *Proc. Natl. Acad. Sci. USA*, 1992, vol. 89, pp. 5814–5818.
24. Smale, S. and Baltimor, D., *Cell*, 1989, vol. 57, pp. 103–113.
25. Klug, A., *Nature*, 1993, vol. 365, pp. 486–487.
26. Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B., *Nature*, 1993, vol. 365, pp. 512–520.
27. Kim, J., Nikolov, D., and Burley, S.K., *Nature*, 1993, vol. 365, pp. 521–527.
28. Coleman, R., Taggart, A., Benjamin, L., and Pugh, B., *J. Biol. Chem.*, 1995, vol. 270, pp. 13842–13849.
29. Wobbe, C.R. and Struhl, K., *Mol. Cell. Biol.*, 1990, vol. 10, pp. 3859–3867.
30. Latchman, D.S., *Eucaryotic Transcription Factors*, New York: Academic, 1995.
31. Buratowski, S., Hahn, S., Guarente, L., and Sharp, P., *Nature*, 1988, vol. 334, pp. 3742–3747.
32. Hoppes, B., Le Blanc, J., and Hawley, D., *J. Biol. Chem.*, 1992, vol. 267, pp. 11539–11547.
33. Imbalzano, A., Zaret, K., Kingsto, R., *J. Biol. Chem.*, 1994, vol. 269, pp. 8280–8286.
34. Hahn, S., Buratowski, S., Sharp P., and Guarente L., *Proc. Natl. Acad. Sci. USA*, 1989, vol. 86, pp. 5718–5722.
35. Coleman, R. and Pugh B.F., *J. Biol. Chem.*, 1995, vol. 270, pp. 13850–13859.

36. Starr, D., Hoopes, B., and Hawley, D., *J. Mol. Biol.*, 1995, vol. 236, pp. 434–446.

37. Suzuki, M., Allen, M., Yagi, N., and Finch, J., *Nucleic Acids Res.*, 1996, vol. 24, pp. 2767–2773.

38. Jou, Z., Chiu, T., Leiberman, P., Baikalov, I., Berk, A., and Dickerson, R., *J. Mol. Biol.*, 1996, vol. 261, pp. 239–254.

39. Sokolenko, A.A., Sandomirskii, I.I., and Savinkova, L.K., *Mol. Biol.*, 1996, vol. 30, pp. 279–285.

40. Bucher, P., *J. Mol. Biol.*, 1990, vol. 212, pp. 563–579.

41. Ponomarenko, M.P. *et al.*, *Computer Analysis of Genetic Macromolecules*, Kolchanov, N.A. and Lim H., Eds., Singapore: World Sci., 1993, pp. 35–67.

42. Kel, A.E. *et al.*, *Comput. Applic. Biosci.*, 1993, vol. 9, pp. 617–627.

43. Likes, J. and Laga, J., *Zakladni statisticke tabulki*, Prague: SNTL, 1978.